



Contribution of rare and predicted pathogenic gene variants to childhood-onset lupus: a large, genetic panel analysis of British and French cohorts

Alexandre Belot, Gillian Rice, Sulliman Ommar Omarjee, Quentin Rouchon, Eve Smith, Marion Moreews, Cécile Frachette, Raphael Bournhonesque, Nicole M. Thielens, Christine Gaboriaud, et al.

► To cite this version:

Alexandre Belot, Gillian Rice, Sulliman Ommar Omarjee, Quentin Rouchon, Eve Smith, et al.. Contribution of rare and predicted pathogenic gene variants to childhood-onset lupus: a large, genetic panel analysis of British and French cohorts. *The Lancet Rheumatology*, 2020, 2 (2), pp.e99-e109. 10.1016/S2665-9913(19)30142-0 . hal-03065973

HAL Id: hal-03065973

<https://hal.science/hal-03065973>

Submitted on 15 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Large-scale analysis of the genetic basis of pediatric systemic lupus erythematosus

Alexandre Belot, MD, PhD^{1, 2, #}, Gillian I Rice, PhD³, Sulliman Ommar Omarjee, MS², Quentin Rouchon, MS^{4, 5}, Eve MD Smith, MD, PhD^{6, 7}, Marion Moreews, MS², Cécile Frachette, MD, MS^{1, 2}, Raphael Bournhonesque, MS², Nicole Thielens, PhD⁸, Christine Gaboriaud, PhD⁸, Isabelle Rouvet, PharmD, PhD⁹, Emilie Chopin, MS⁹, Bruno Ranchin, MD¹, Rolando Cimaz, MD¹⁰, Paula Romagnani, MD, PhD¹¹, Christophe Malcus, PharmD, PhD¹², Nicole Fabien, PharmD, PhD¹³, Marie Nathalie Sarda, MD, PhD¹³, Behrouz Kassai, MD, PhD¹⁴, Jean-Christophe Lega, MD, PhD¹⁵, Stéphane Decramer, MD, PhD¹⁶, Claire Bardel, PhD^{17, 18}, Pierre Antoine Rollat-Farnier PhD¹⁸, Sebastien Viel, PharmD, PhD¹³, Héloïse Reumaux, MD¹⁹, James O'Sullivan, MS²⁰, Jamie Ellingford, PhD²⁰, Tracy A. Briggs, MD, PhD^{3, 20}, Thierry Walzer, PhD², Anne-Laure Mathieu, PhD², Brigitte Bader-Meunier MD²¹, Michael W Beresford, MD, PhD^{6, 7}, Yanick J Crow MD, PhD^{22, 23, 24, #}, GENIAL Study Group, UK JSLE Study Group.

1. Pediatric Nephrology, Rheumatology, Dermatology Unit, Femme Mere Enfant Hospital, Hospices Civils de Lyon, France
2. CIRI, Centre International de Recherche en Infectiologie / International Center for Infectiology Research, Inserm, U1111, Ecole Normale Supérieure de Lyon, Université Lyon 1, CNRS, UMR5308, Lyon, France
3. Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, M13 9PT, United Kingdom
4. Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium
5. Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium
6. Paediatric Rheumatology, Alder Hey Children's NHS Foundation Trust, Liverpool, United Kingdom
7. Department of Women and Children's Health, Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom
8. University of Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France
9. Centre de biotechnologie cellulaire et Biothèque, Groupe Hospitalier Est, Hospices Civils de Lyon, 69677 Bron, France
10. Rheumatology Unit, Anna Meyer Children Hospital and University of Florence, University of Florence, Florence, Italy
11. Nephrology Unit, Anna Meyer Children Hospital and University of Florence, University of Florence, Florence, Italy

12. Service d'Immunologie, Hôpital Edouard Herriot, Hospices Civils de Lyon, 69437 Lyon, France
13. Service d'immunologie, Hospices Civils de Lyon, CHLS, 69495 Pierre-Bénite, France
- 14: EPICIME-CIC 1407 de Lyon, Inserm, Service de Pharmacotoxicologie, Hospices Civils de Lyon & Université Lyon 1, 69677, Bron, France
15. Internal Medicine Unit, CHLS, Hospices Civils de Lyon, Pierre Benite, Université de Lyon 1, Lyon, France
16. Department of Pediatric Nephrology, Centre Hospitalier Universitaire de Toulouse, Toulouse, France. Centre De Référence des Maladies Rénales Rares du Sud Ouest & Inserm U1048, France
17. Department of biostatistics – bioinformatics, Hospices Civils de Lyon, 69677, Bron, France
18. CNRS UMR5558, Biometry and evolutionaty biology lab, Lyon University, Lyon 1 University, F-69622 Villeurbanne, France
19. CHU Lille, pediatric rheumatology unit, 59000 Lille, France
20. Manchester Centre for Genomic Medicine, Manchester Academic Health Sciences Centre, Manchester University NHS Foundation Trust, St Mary's Hospital, Manchester, United Kingdom
21. Pediatric Rheumatology and Immunology Unit, Necker hospital, Imagine Institution, Paris, France
23. Laboratory of Neurogenetics and Neuroinflammation, Institut Imagine, Paris, France
24. Paris Descartes University, Sorbonne-Paris-Cité, Institut Imagine, Paris, France
25. Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

Collaborators (21)

From GENIAL Study Group: Aboujaoude P, Bouayed K, Burtsey S, Despert V, Fain O, Fischbach M, Hachulla E, Hatchuel Y, Kone-Paut I, Maurier F, Meinzer U, Piram M

From UK JSLE Study Group: Al-Abadi E, Ciurtin C, Gardner-Medwin J, Leone V, McErlane F, Pilkington C, Riley P, Satyapal R, Sridhar A.

Joint corresponding authors (Alexandre Belot, alexandre.belot@chu-lyon.fr and Yanick Crow, yanick.crow@mac.com)

Large-scale analysis of the genetic basis of pediatric systemic lupus erythematosus

Abstract

Background

Systemic lupus erythematosus (SLE) is a rare immunological disorder where genetic factors are important in causation. Mendelian forms of lupus have been described in the context of almost 30 genotypes in humans, and more than 60 in mice. Murine susceptibility models and genome-wide association studies (GWAS) also highlight the role of genetic variants in pathogenesis. The overall genetic contribution to pediatric SLE is unknown.

Methods

We designed a next-generation sequencing panel comprising 147 genes, including all known Mendelian lupus causing (KLC) genes in humans, and lupus associated genes identified through GWAS and animal models (potentially lupus causing, PLC, genes). Using this panel we screened 117 probands fulfilling American College of Rheumatology criteria for SLE, ascertained through two cohorts of pediatric SLE in the UK and France, and 791 ethnically matched controls from the 1000 Genomes Project.

Results

Mendelian genotypes were present in 6.8% of probands. Beyond these cases, rare, predicted damaging variants were significantly enriched in the SLE cohort compared to controls, with an odds ratio of 14.09 and 3.99 in KLC and PLC genes respectively. Overall, 27% of SLE probands versus 4.6% of controls were identified with at least one rare, predicted damaging variant amongst our selected gene panel ($p = 4.14 \times 10^{-15}$).

Conclusion

Rare and predicted damaging variants in KLC and PLC genes were highly enriched in a population of pediatric onset lupus, with 1 in 15 probands demonstrating clear Mendelian causation. Germline defects of innate immunity represent the main genetic contribution to SLE in children.

Word count: 2733

Introduction

The term systemic lupus erythematosus (SLE) describes a rare, heterogeneous set of phenotypes characterized by the presence of autoantibodies targeting nuclear autoantigens and type I interferon upregulation. Familial aggregation and higher concordance rates between monozygotic compared to dizygotic twins suggest a major hereditary component to pathogenesis^{1,2}. Genome-wide association studies (GWAS) have identified more than 80 SLE-associated loci^{3,4}. GWAS defined variants are common, confer small effects on disease susceptibility and typically fall outside of coding regions⁵. In contrast, monogenic forms of SLE, of which almost 30 have been described in humans, and greater than 60 in mice⁶, involve highly penetrant rare variants in protein encoding DNA⁷.

Mendelian forms of disease can help define the involvement of discrete pathways in pathogenesis. For example, both classical complement gene mutations and DNASE1L3 deficiency highlight the importance of efferocytosis in lupus pathology^{8,9}, whilst the relevance of type I interferon signaling to SLE is underlined by the association with the Mendelian type I interferonopathies^{10–13}. Finally, B cells are a key player in lupus causation¹⁴, with PKC δ deficiency the first described B cell related form of monogenic lupus¹⁵, and heterozygous mutations in *IKZF1*, encoding IKAROS, a B cell transcription factor, a cause of germline autoimmunity including SLE^{16,17}.

Recent sequencing studies suggest a false dichotomy in categorizing diseases according to strict complex or Mendelian models, an alternative possibility highlighting the importance of combinations of small numbers of rare variants promoting disease in a single individual and at a population level^{18,19}. Immune responses are variable in humans, with up to 40% of this diversity estimated to be explained by genetic variation²⁰. In association with environmental factors, such genetic polymorphisms may promote tolerance breakdown²¹, as exemplified by the lupus phenotype.

Here, we describe a large-scale molecular analysis of pediatric SLE, highlighting a mutational spectrum directly causing, or strongly influencing, pathogenesis.

Material and methods

Patients

We identified 117 unrelated probands from the UK Juvenile-onset SLE (JSLE)²² and French SLE GENIAL cohorts fulfilling American College of Rheumatology (ACR) criteria for lupus, diagnosed before the age of 16 years. The study was approved by the relevant ethics committees (see the Supplementary Appendix for further details). The combined cohort comprised the following ethnicities: European (50%), African (33%) and South Asian (16%), classified according to International Genome Sample resource criteria (Figure 1A).

Gene selection

To identify genetic variations potentially responsible for SLE, we selected a set of genes based on: i) known Mendelian forms of lupus ('known lupus causing' – KLC - genes; $n = 28$), some also detected by GWAS and/or in murine studies (Table S1 in the Supplementary Appendix); ii) defined causes of monogenic lupus in mouse models ($n = 62$) and/or genes demonstrating an association with SLE identified through GWAS ($n = 67$), with 11 genes common to both. The latter group are referred to as 'putative lupus causing' (PLC) genes (Table S2 in the Supplementary Appendix). Defined as of 2013 when the study was initiated, this gene set comprised 147 in total.

DNA sequencing, variant discovery and filtering

Details of our sequencing protocols and variant analyses are provided in the Supplementary Appendix. In brief, targeted enrichment and sequencing of the 147 gene panel was undertaken using DNA extracted from peripheral blood. Single nucleotide variants (SNV) and indels were identified according to strict bioinformatic protocols, aligned with 1000 Genomes Project (1000g)²³ sequence sets, and variant pathogenicity assessed using SIFT²⁴, Polyphen-2²⁵ and Combined Annotation Dependent Depletion (CADD) v1.3²⁶ scores. SNVs were filtered according to their population frequencies within the Exome Aggregation Consortium (ExAC) dataset, employing 1% and 0.1% in the case of putative biallelic mutations, or 0.01% and 0.001% for heterozygous variants in the case of indels or SNVs respectively, as our thresholds (threshold (heterozygous) = threshold (homozygous)² i.e. 10^{-4} for indels and 10^{-6} for SNVs). We selected indels with a CADD score higher than 15, and SNVs according to the formula: CADD score ≥ 15 AND sift score ≤ 0.05 AND Polyphen 2 score ≥ 0.9 . A cartoon of the characterization of different sequence variants is given in Figure S1 in the Supplementary Appendix.

Control cohort

Control data were derived from the 1000 Genomes Project (1000g)²³. The healthy control cohort comprised 791 unrelated individuals demonstrating a similar ethnic distribution to the SLE cohort (50% European, 33% African; 16% South Asian), built by random sampling among the 1000g sample set.

***In vitro* assays, structural and network analyses**

Details of *in vitro* experiments, structural studies and protein interaction network construction are given in the Supplementary Appendix.

Statistical analysis

Fischer exact testing and Bonferroni correction were used to compare the frequencies of variants in the control and lupus cohorts. Data normality was assessed using Shapiro-Wilk test and quantile-quantile plots. The mean of SNPs per individual was compared between control and lupus cohorts using a Mann-Whitney U test (Wilcoxon rank-sum test), and a Bonferroni correction applied.

Results

Population characteristics

Our patient cohort demonstrated a median age at disease onset of 12 (range 1.8 - 16) years and sex ratio of 25/92 (male:female = 1:3.7) (Figure 1B). Similar to previous studies of childhood SLE, major organ involvement was frequent, with renal and cerebral disease noted in 59% and 23% of individuals respectively. A family history of lupus was reported in 20% of cases.

Next generation sequencing identified a total of 97,855 variants in the 147 KLC and PLC genes across 117 patients (Figures 1C, 1D). Filtering according to the strategy described above was used to select for very rare variants, allowing us to exclude 99.95% of these initially identified polymorphisms. A similar strategy was applied to the control cohort of 791 ethnically matched individuals (Figure 1E).

Prevalence of rare and predicted damaging variants in the 28 KLC genes

After filtering, 16 unrelated probands were identified to harbor a total of 15 either heterozygous or biallelic variants in 12 KLC genes (Table 1), relating to components of the classical complement pathway (*C1QA*, *C1QC*, *C2*, *C3*, *C5*, *C8B*, *C9*), DNASE1L3 deficiency (*DNASE1L3*), type I interferonopathies (*TREX1*, *RNASEH2C*, *ACP5*) or B-cell dysfunction (*IKZF1*).

Segregation was compatible with autosomal recessive inheritance due to complement deficiency in six pedigrees. Biallelic mutations in either *C1QC* or *C1QA* were identified in three families (Figure 2A). *In silico* structural analysis of *C1qC* predicted that the p.Gly164Ser substitution, already reported as pathogenic²⁷, destabilizes the globular head of the *C1q* complement protein (Figure 2B). Two other variants in *C1QC* and *C1QA* were nonsense mutations. *C1q* was undetectable in patient 1 (Pt1) (Figure 2C). Patients 4 to 6 carried the same biallelic deletion in the *C2* gene (Figure 2D), previously described as a cause of lupus with incomplete penetrance²⁸. CH50 and *C2* levels were decreased in Pt6, whilst *C3* and *C4* were normal (Figure 2E). Two predicted damaging variants in *DNASE1L3* were identified in Pt7, and biallelic inheritance confirmed by parental testing (Figure 2F). The nonsense mutation has been reported as disease causing²⁹, whilst the deletion of exon 5 was novel, located in the deoxyribonuclease domain (Figure 2G). Sequencing of the family of Pt8 demonstrated segregation of a novel missense *IKZF1* variant with disease status in two additional relatives (Figure 2H). A previous GWAS identified significant linkage at the *IKZF1* locus with lupus³⁰, whilst loss-of-function germline mutations have been reported in association with systemic autoimmunity and B cell deficiency^{16,17}. The p.Asp120Val *IKZF1* variant is located in the first zinc finger of the protein (Figure 2I). Confocal microscopy of NIH-3T3 cells transfected with wild type (WT) and mutant constructs revealed a punctate staining pattern characteristic of pericentromeric heterochromatin binding and localization of the WT protein, whereas the mutant protein exhibited diffuse nuclear staining.

Eight additional probands carried a monoallelic, rare, predicted damaging variant in a KLC gene previously described to cause SLE as an autosomal recessive trait (Table 1). A *TREX1* p.Ser82Leufs*9 frameshift mutation has been reported as disease-causing in Aicardi-Goutières syndrome (AGS)³¹. Pt16, carrying a *C3* substitution, demonstrated early onset lupus nephritis with *C1q* deposits and thrombotic microangiopathy. *C3* was undetectable on several assessments. Heterozygous variants in *C3* have been reported as possibly causal in SLE³². Other predicted damaging variants were recorded in *RNASEH2C*, *ACP5*, *C5*, *C8B* and *C9*. In this group of eight patients the sex ratio was 1:1, similar to the subset of probands described above, but differing from the overall composition of the cohort (Figure S2 in the Supplementary Appendix), and the median age at onset was earlier than in the cohort overall.

To test if the accumulation of heterozygous KLC variants was specifically associated with lupus status, we applied the same algorithm to our control group as used in the lupus data set. In so doing we observed a statistically significant excess of rare, predicted damaging variants in KLC genes within the pediatric SLE population compared to controls, with an odds ratio of 14.10 in the SLE cohort (Figure 3A). These data comprised all variants in our gene panel in the SLE and 1000g cohort datasets, generated respectively by targeted and whole genome sequencing. Considering that sequence coverage and indel analysis could be different between our in-house derived sequence and those of the 1000g, we carried out the similar assessment but restricted to SNVs within the coding regions of the KLC gene set (Figure S1, Figure S3A and Supplemental table 3 in the Supplementary Appendix). Again, rare and predicted damaging variants in KLC genes were significantly increased in the SLE population ($p = 4.11 \times 10^{-4}$).

Prevalence of rare and predicted damaging variants in 119 PLC genes

Focusing on variants in PLC genes, we detected 19 rare, predicted damaging PLC variants in 16 additional unrelated probands (Table 2). Two patients carried the same novel substitution

in *PDHX*. In both cases there was a family history of SLE consistent with autosomal dominant inheritance. Samples were unavailable from other relatives so that we could not assess segregation. *PDHX*, included in our gene panel because of its identification in a GWAS of SLE³³, encodes pyruvate dehydrogenase, a key glycolytic enzyme whose expression is reduced in white blood cells of lupus patients³⁴.

As for the KLC gene set, we observed a statistically significant enrichment of PLC filtered variants in the lupus population compared to the 1000g controls, with an odds ratio of 3.99 (Figure 3B) and an increase in the number of filtered SNVs per individual in the coding portion of our PLC gene set compared to controls ($p = 6.4 \times 10^{-3}$) (Figure S3B in the Supplementary Appendix).

Finally, we compared the combined frequency of KLC and PLC rare and predicted damaging variants per proband in the two cohorts, and identified a statistically significant increase of one, two and three genes with filtered variants per individual in the SLE population (Figure 3C, 3D).

Phenotype according to gene-related pathway

To determine whether genotype might influence phenotype, we considered the 30 genes with at least one rare, predicted damaging variant in our lupus cohort and assessed the protein interaction network built around these genes using String software³⁵. Most of the genes were connected, defining a functional lupus interactome (Figure 4A). We then analyzed features of patients represented within one of three pathophysiological categories (efferocytosis, type I interferon signaling and primary adaptive immunity). The pattern of clinical expression was similar across the first two pathways. In contrast, patients with putative defects in adaptive immunity demonstrated renal and musculoskeletal involvement less frequently (Figures 4B - D).

Discussion

Immunoprofiling of pediatric SLE patients has highlighted the heterogeneous nature of this phenotype³⁶. Whilst the basis of such heterogeneity remains poorly defined, the identification of Mendelian forms of SLE related to different pathogenic mechanisms indicates that genetic factors are likely important. Low frequency variants are not captured by GWAS, nor do they confer sufficiently large effect sizes to be detected by classical linkage analysis in small family studies⁷. Meanwhile, the observation of variable penetrance and expression in Mendelian disorders indicates that other genes or environmental factors can impact phenotype.

A major challenge of our study was to set appropriate thresholds, in terms of population frequency and pathogenicity prediction, in order to capture likely causal variants. Given the identification of unequivocal Mendelian genotypes, the algorithms we employed appear valid. This bioinformatics strategy also demonstrated that the distribution of filtered variants was not equal in ethnically matched patient and control populations. Such stringent filtering, leading to an enrichment for likely pathogenic alleles, risks excluding possible disease-causing variants. For example, a p.Asp105Ala heterozygous variant in *RNASEH2B*, previously considered as causal for SLE¹¹, was excluded using this filtering strategy (data not shown). As such, our results likely represent a conservative estimate of the associated high-penetrant genetic load.

Our analysis identified eight probands with mutations in KLC genes consistent with Mendelian causation. Complement deficiencies accounted for 5% of probands in our cohort and a defect of efferocytosis is also relevant in *DNASE1L3* deficiency⁹. Only one patient in our cohort was identified with a monogenic disease involving B cell dysfunction. That is, a novel mutation in *IKZF1*, encoding the transcription factor IKAROS, playing a role in B cell development. No mutations were identified in *PRKCD*, previously associated with monogenic lupus¹⁵.

In lupus patients compared to controls, we observed an increased frequency of monoallelic variants in KLC genes described to cause lupus as an autosomal recessive trait in both type I interferonopathy and complement related genes. In this group the sex ratio was 1:1, and the onset of disease was very early, possibly suggesting a significant impact of the observed heterozygous variants on disease induction. Among these variants, a TREX1 mutation has already been reported in AGS, a recognized type I interferonopathy associated with an increased risk of SLE¹⁰. Günther et al. also described an increase of rare heterozygous variants in components of the RNase H2 complex in lupus¹¹. Using our filtering strategy, one patient was detected with an RNASEH2C monoallelic, predicted pathogenic substitution. We recorded an *ACP5* missense variant in one patient. Biallelic mutations in *ACP5* are a cause of Mendelian lupus¹², and an increase of predicted pathogenic heterozygous variants in this gene has been reported in an adult lupus cohort¹³. The patient carrying the novel variant in *C3* displayed features of lupus nephritis and hemolytic uremic syndrome (HUS), with *C3* deficiency a known cause of HUS. Overall, these discoveries support the possibility that genes known as directly causal for Mendelian lupus may also play a role in driving susceptibility to complex disease. We note that *de novo* monoallelic mutations have recently been described in KLC genes in lupus patients, supporting the contention of a contribution of heterozygous variants to SLE susceptibility³⁷.

We identified 16 patients with heterozygous, putative mutations in PLC genes. The frequency of these rare, predicted damaging variants was significantly increased in lupus patients compared to controls, suggesting that they might represent novel SLE predisposing alleles. Furthermore, some filtered variants co-segregated in a single individual. For example, Pt8 carried a *LYN* substitution associated with the novel *IKZF1* mutation discussed above, whilst Pt10 was identified with a novel, predicted damaging variant in each of *ACP5* and *RASGRP3*, and a very rare variant in *FCGR3A*. It seems plausible that a lupus phenotype might be driven by pathogenic variants in two or more genes.

Using our strategy, focusing on exons and polymorphisms with predicted functional consequences, we identified putative disease causing variants in coding regions of genes already detected in GWAS studies. As examples, we recorded two unrelated patients carrying the same variant in *PDHX*, and a novel *PTPN22* variant was observed in one patient. Furthermore, three patients carried the identical, very rare indel in the *FCGR3A* gene encoding Fc gamma receptor III. Rare variants may contribute to the burden of the disease, with their accumulation influencing age at onset and clinical spectrum.

Studying extreme phenotypes of complex disease represents a powerful strategy to simplify and understand human pathology. Here, we highlight at least three distinct immune defects driving pathogenesis in pediatric SLE, with the majority of disease-associated alleles detected in genes involved in innate immunity (Figure S4). Supporting the conclusions of a recent transcriptomic study³⁶, our analysis highlights the heterogeneity of the genetic basis of pediatric SLE, with 27% of the pediatric SLE population carrying at least one rare and predicted damaging variant. Numerous clinical trials have failed to demonstrate a positive effect of a variety of medications in SLE³⁸, and current clinical and laboratory criteria are clearly not sufficient to detect specific factors determining treatment outcome³⁹. Explicitly considering rare, high penetrant gene variants might identify patient subsets with distinct therapeutic responses, thereby enabling personalized treatments according to genetic background and molecular taxonomy⁴⁰.

Acknowledgements

YJC acknowledges the European Research Council (GA 309449; and 786142 - E-T1IFNs), ERA-NET Neuron (MR/M501803/1) and a state subsidy managed by the National Research Agency (France) under the "Investments for the Future" program bearing the reference ANR-10-IAHU-01. AB acknowledges the French ANR program 14-CE14-0026 LUMUGENE and the

“Jeune Chercheur” support from Hospices Civils de Lyon. TAB acknowledges the National Institute for Health Research (NIHR) (NIHR Transitional Research Fellowship, Dr Tracy Briggs, TRF-2016-09-002). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. J.M.E. is funded by a postdoctoral research fellowship from the Health Education England Genomics Education Programme (HEE GEP). The views expressed in this publication are those of the authors and not necessarily those of the HEE GEP. MWB acknowledges the UK JSLE Cohort Study is supported by Lupus UK (#LUPUS UK: JXR10500, JXR12309], the UK’s ‘Experimental Arthritis Treatment Centre for Children’ supported by Arthritis Research UK (ARUK #20621), the Alder Hey Charity, University of Liverpool and Alder Hey Children’s NHS Foundation Trust, and the National Institute of Health Research (NIHR) Alder Hey Clinical Research Facility. The authors would like to acknowledge all patients and their families for participating in UK JSLE Cohort Study and in GENIAL cohort study and are grateful to all the support given by the local principle investigators and entire multi-disciplinary teams within each of the paediatric centres as part of the UK JSLE Study Group (<https://www.liverpool.ac.uk/translational-medicine/research/ukjsle/jsle/>) and GENIAL Study Group.

References

1. Criswell LA. The genetic contribution to systemic lupus erythematosus. *Bull NYU Hosp Jt Dis* 2008;66(3):176–83.
2. Crow YJ. Lupus: how much “complexity” is really (just) genetic heterogeneity? *Arthritis Rheum* 2011;63(12):3661–4.
3. Bentham J, Morris DL, Graham DSC, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* 2015;47(12):1457–64.
4. Morris DL, Sheng Y, Zhang Y, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet* 2016;48(8):940–6.
5. Chen L, Morris DL, Vyse TJ. Genetic advances in systemic lupus erythematosus: an update. *Curr Opin Rheumatol* 2017;29(5):423–433.
6. Kono DH, Theofilopoulos AN. Genetics of SLE in mice. *Springer Semin Immunopathol* 2006;28(2):83–96.
7. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461(7265):747–53.
8. Al-Mayouf SM, Sunker A, Abdwani R, et al. Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. *Nat Genet* 2011;43(12):1186–8.
9. Sisirak V, Sally B, D’Agati V, et al. Digestion of Chromatin in Apoptotic Cell Microparticles Prevents Autoimmunity. *Cell* 2016;166(1):88–101.
10. Lee-Kirsch MA, Gong M, Chowdhury D, et al. Mutations in the gene encoding the 3’-5’ DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat Genet* 2007;39(9):1065.
11. Günther C, Kind B, Reijns MAM, et al. Defective removal of ribonucleotides from DNA promotes systemic autoimmunity. *J Clin Invest* 2015;125(1):413–24.
12. Briggs TA, Rice GI, Daly S, et al. Tartrate-resistant acid phosphatase deficiency causes a bone dysplasia with autoimmunity and a type I interferon expression signature. *Nat Genet* 2011;43(2):127–31.
13. An J, Briggs TA, Dumax-Vorzet A, et al. Tartrate-Resistant Acid Phosphatase Deficiency in the Predisposition to Systemic Lupus Erythematosus. *Arthritis Rheumatol Hoboken NJ* 2017;69(1):131–42.
14. Lipsky PE. Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. *Nat Immunol* 2001;2(9):764.
15. Belot A, Kasher PR, Trotter EW, et al. Protein kinase cδ deficiency causes mendelian systemic lupus erythematosus with B cell-defective apoptosis and hyperproliferation. *Arthritis Rheum* 2013;65(8):2161–71.

16. Hoshino A, Okada S, Yoshida K, et al. Abnormal hematopoiesis and autoimmunity in human subjects with germline IKZF1 mutations. *J Allergy Clin Immunol* 2017;140(1):223–31.
17. Kuehn HS, Boisson B, Cunningham-Rundles C, et al. Loss of B Cells in Patients with Heterozygous Mutations in IKAROS. *N Engl J Med* 2016;374(11):1032–43.
18. Jordan DM, Do R. Using Full Genomic Information to Predict Disease: Breaking Down the Barriers Between Complex and Mendelian Diseases. *Annu Rev Genomics Hum Genet* 2018;19:289–301.
19. Sazonovs A, Barrett JC. Rare-Variant Studies to Complement Genome-Wide Association Studies. *Annu Rev Genomics Hum Genet* 2018;19:97–112.
20. Brodin P, Davis MM. Human immune system variation. *Nat Rev Immunol* 2017;17(1):21–9.
21. Davis MM, Brodin P. Rebooting Human Immunology. *Annu Rev Immunol* 2018;36:843–64.
22. Watson L, Leone V, Pilkington C, et al. Disease activity, severity, and damage in the UK juvenile-onset systemic lupus erythematosus cohort. *Arthritis Rheum* 2012;64(7):2356–65.
23. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061.
24. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31(13):3812–4.
25. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013;Chapter 7:Unit7.20.
26. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46(3):310–5.
27. López-Lera A, Torres-Canizales JM, Garrido S, Morales A, López-Trascasa M. Rothmund-Thomson syndrome and glomerulonephritis in a homozygous C1q-deficient patient due to a Gly164Ser C1qC mutation. *J Invest Dermatol* 2014;134(4):1152–4.
28. Johnson CA, Densen P, Wetsel RA, Cole FS, Goeken NE, Colten HR. Molecular heterogeneity of C2 deficiency. *N Engl J Med* 1992;326(13):871–4.
29. Özçakar ZB, Foster J, Diaz-Horta O, et al. DNASE1L3 mutations in hypocomplementemic urticarial vasculitis syndrome. *Arthritis Rheum* 2013;65(8):2183–9.
30. Han J-W, Zheng H-F, Cui Y, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* 2009;41(11):1234–7.
31. Crow YJ, Chase DS, Lowenstein Schmidt J, et al. Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *Am J Med Genet A* 2015;167A(2):296–312.
32. Nozal P, Garrido S, Martínez-Ara J, et al. Case report: lupus nephritis with autoantibodies to complement alternative pathway proteins and C3 gene mutation. *BMC Nephrol* 2015;16:40.
33. Lessard CJ, Adrianto I, Kelly JA, et al. Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study. *Am J Hum Genet* 2011;88(1):83–91.
34. Lee H-T, Lin C-S, Lee C-S, Tsai C-Y, Wei Y-H. Increased 8-hydroxy-2'-deoxyguanosine in plasma and decreased mRNA expression of human 8-oxoguanine DNA glycosylase 1, anti-oxidant enzymes, mitochondrial biogenesis-related proteins and glycolytic enzymes in leucocytes in patients with systemic lupus erythematosus. *Clin Exp Immunol* 2014;176(1):66–77.
35. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;43(Database issue):D447–452.
36. Banchereau R, Hong S, Cantarel B, et al. Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell* 2016;165(6):1548–50.
37. Pullabhatla V, Roberts AL, Lewis MJ, et al. De novo mutations implicate novel genes in systemic lupus erythematosus. *Hum Mol Genet* 2018;27(3):421–9.

38. Isenberg DA, Merrill JT. Why, why, why de-lupus (does so badly in clinical trials). *Expert Rev Clin Immunol* 2016;12(2):95–8.
39. Weill O, Decramer S, Malcus C, et al. Familial and syndromic lupus share the same phenotype as other early-onset forms of lupus. *Jt Bone Spine Rev Rhum* 2016;84:589-93.
40. Barturen G, Beretta L, Cervera R, Van Vollenhoven R, Alarcón-Riquelme ME. Moving towards a molecular taxonomy of autoimmune rheumatic diseases. *Nat Rev Rheumatol* 2018;14(3):180.

Figure 1. Demographic characteristics of the pediatric SLE cohort and bioinformatic analysis strategy. Ethnicity (A), age and sex (B) of the 117 probands in the pediatric SLE cohort. (C) Schematic of the bioinformatic pipeline used to detect rare predicted damaging variants, leading to the definition of three categories of patient: 8 patients with monogenic SLE; 8 patients carrying monoallelic variants in autosomal recessive known lupus causing (KLC) genes; 16 additional patients with filtered variants in potentially lupus causing (PLC) genes. Total and individual variant numbers are presented before and after the application of filters of rarity and pathogenicity in the SLE (D) and 1000g (E) cohorts. Note, three patients in the KLC groups carried additional rare and predicted damaging variants in PLC genes, as indicated in brackets.

Figure 2. Eight monogenic forms of pediatric SLE. Pedigrees of families with mutations in C1q (A), C2 (D), DNASE1L3 (F) and IKAROS (H) (black shading represents affected SLE status), with identified mutations shown below each pedigree. (B) Structure of C1q globular domain with C1qA (red), C1qB (green) and C1qC (blue), calcium ion (yellow) and representation of the G164 residue (magenta). This amino acid is highly conserved across the C1q family. The glycine at position 164 is located near the junction between the collagen-like stem (red, green and blue circles) with the globular domain. The glycine is tightly packed into the structure, and any substitution is predicted to destabilize the assembly of the three chains. (C) ELISA quantification of complement components in sera from patients with C1Q and C2 deficiency (E). (G) Representation of the DNASE1L3 protein showing the T97Ifs*2 substitution and the deletion identified in Pt7 (red), and the only other mutation so-far reported within the coding region (black). (I) Position of the D120V mutation (red) in the IKAROS zinc finger domain required for DNA-binding, and previously reported mutations (black). NIH3T3 cells were transfected with HA-tagged WT or D120V mutant expression vectors labeled with anti-HA antibody and an Alexa 488-conjugated (green) secondary antibody. Cells were visualized using confocal microscopy.

Figure 3. Bioinformatic filtering of rare and predicted damaging variants in pediatric SLE patients and controls. (A) Heterozygous variant enrichment in known lupus causing (KLC) genes (indels and SNVs), previously associated with lupus causation as an autosomal recessive trait, in SLE probands compared to controls. Note, the eight SLE patients considered as having a proven Monogenic cause for their disease (see Figure 2) have been excluded from this analysis so as to avoid biasing the data. (B) Total filtered heterozygous variant enrichment in possibly lupus causing (PLC) genes (indels and SNVs) in SLE probands compared to controls. (C) Number of KLC and PLC genes per individual in which a rare, predicted damaging variant was identified in pediatric SLE probands and controls. (D) Representation of genes with either filtered variants (inner white circle) or without filtered variants (outer grey circle). Chromosome localization and number of filtered variants (innermost colored dots) are indicated. Variants were divided into three categories: causal for monogenic disease (red), heterozygous in KLC genes previously associated with lupus causation as an autosomal recessive trait (orange) or PLC (blue) genes. Putative di- or tri-genic associations of variants in a single proband are signified by connecting lines.

Figure 4. Phenotype according to genetic background. Interactome of lupus susceptibility genes constructed using String software (version 9.0; <http://string-db.org>) including all KLC and PLC genes in which at least one rare, predicted damaging variant (n = 30) was identified in the lupus population; solid lines highlight the main pathways (A). Spider plots of the frequency of clinical features, and dsDNA positivity observed in pediatric SLE probands carrying filtered variants in efferycytosis related genes (B), type I interferon related genes (C), and B cell related genes (D).

Supplementary figure 1. Coding region filter. The red box outlines what we refer to as the coding region i.e. exons and canonical splice site nucleotides.

Supplementary figure 2. Phenotype according to genotype. Spider plots of the frequency of clinical features seen in: A) lupus patients with Mendelian disease (n = 8); (B) probands carrying heterozygous variants in autosomal recessive KLC genes (n = 8); (C) probands with variants in PLC genes (n = 16); and (D) probands with no identified filtered variants (n = 85). Sex ratios and age at diagnosis are provided for each plot. Sex ratio, male:female; Age, median age at onset in years; SD: standard deviation.

Supplementary figure 3. Bioinformatic filtering of rare and predicted damaging variants in pediatric SLE patients and controls restricted to single nucleotide variants (SNVs) in coding regions. Variant number of KLC (A) or PLC (B) heterozygous (het) SNVs in coding regions per individual in SLE probands and controls. Note, the eight SLE patients considered as having a proven Monogenic cause for their disease (see Figure 2) have been excluded from this analysis so as to avoid biasing the data.

Supplementary figure 4. Model of the relationship of identified rare and predicted damaging variants to pediatric lupus pathogenesis. This model highlights the pathways identified by the current study considering rare, predicted damaging variants identified in pediatric SLE probands. Processing and clearance of immune complexes and interferon signaling are represented in the upper panel; signal transduction in the adaptive immune response is represented in the lower panel. Abbreviations: AB, apoptotic bodies; IFN, interferon; MΦ, macrophage; pDC, plasmacytoid dendritic cells.

Table 1: Patients carrying filtered variant in known lupus causing (KLC) genes (bold) (with additional filtered variants, underlined), defining Mendelian SLE (Pt 1 to 8) or carrying single rare, predicted damaging heterozygous variants in genes known to cause autosomal recessive lupus (Pt9 to 16)

Patient	Homozygous variants	Compound heterozygous variants	Single heterozygosity	SIFT	Polyphen2	CADD	ExAc	References
Pt1	C1QC c.490G>A p.Gly164Ser			Deleterious	Probably damaging	21.3	1/121376	
Pt2		C1QC c.120delC p.Leu41Cysfs*97 / c.490G>A p.Gly164Ser		-	-		Novel	27
			<u>CLEC16A</u> in-frame deletion g:11073185 G/-	Deleterious	Probably damaging	32	1/121376	
				-	-	15.75	Novel	
Pt3	C1QA c.44delT p.Ile15Asnfs*7 C2	No	No	-	-	21.3	Novel	
Pt4	c.839_849+17delTGGT GGACAGGGTCAGGAA TCAGGAGTC			-	-	32	619/121204 1 hom	
Pt5	C2 c.839_849+17delTGGT GGACAGGGTCAGGAA TCAGGAGTC			-	-	32	619/121204 1 hom	28
Pt6	C2 c.839_849+17delTGGT GGACAGGGTCAGGAA TCAGGAGTC			-	-	32	619/121204 1 hom	
Pt7		DNASE1L3 c.290_291delCA p.Thr97IlefsTer2 / Exon 5 deletion		-	-	25.1	10/121384	29
				-	-		Novel	
Pt8			IKZF1 c.199A>T p.Ter67Cys <u>LYN</u> c.773T>C p.Phe258Ser	Deleterious	Probably damaging	26.3	Novel	
				Deleterious	Probably damaging	24.2	Novel	

Pt9	TREX1 c.226_227insCCTG CAGC het p.Ser82Leufs*9	-	-	24.6	Novel
Pt10	ACP5 c.79C>A p.Arg27Ser	Deleterious	Probably damaging	28.1	Novel
	RASGRP3 c.686C>G p.Ala229Gly	Deleterious	Probably damaging	31	Novel
FCGR3A intronic variant: g.161514759 delC ins CAG					
Pt11	RNASEH2C c.472C>G p.His158Asp	Deleterious	Probably damaging	28.8	Novel
Pt12	C9 c.1209_1222delTGAT TGTGTAAAGA p.Asp403GlufsTer5	-	-	34	8/121256
Pt13	C8B p.Ala4STOP (indel)	-	-	16.31	Novel
Pt14	C5 p.Asp966Tyr	-	-	16.5	1/30882
Pt15	C9 c.1490C>T p.Ala497Val	Deleterious	Probably damaging	29.2	Novel
Pt16	C3 c.3278T>G p.Ile1093Ser	Deleterious	Probably damaging	31	Novel
Pt11	RNASEH2C c.472C>G p.His158Asp	Deleterious	Probably damaging	28.8	Novel
Pt12	C9 c.1209_1222delTG	-	-	34	8/121256

Pt13	ATTGTGTAAAGA p.Asp403GlufsTer5 C8B p.Ala4STOP (indel)	-	-	16.31	Novel
Pt14	C5 p.Asp966Tyr	-	-		
Pt15	C9 c.1490C>T p.Ala497Val	Deleterious	Probably damaging		
Pt16	C3 c.3278T>G p.Ile1093Ser	Deleterious	Probably damaging		

Table 2: Rare, predicted damaging variants per patient in putative lupus causing (PLC) genes (GWAS and mouse model), Pt 17 to 32

Patient	Single heterozygosity	SIFT	Polyphen2	CADD	ExAc
Pt17	TNIP1 g.10444711, delA	-	-	18.67	Novel
Pt18	IRF7 c.926A>T p.Lys309Met	Deleterious	Probably damaging	24.9	Novel
Pt19	PXK g.583955125 delAAGTAGACAGCCTGAGC	-	-	15.65	Novel
Pt20	FCGR3A g.161514759 delC ins CAG	-	-	18.07	1/114654
Pt21	CLEC16A c.904_906delGGA p.Gly302del	-	-	15.52	Novel
	NFE2L2 c.- 11_9delCACAGCTCATCATGATGGAC -	-	-	22.8	Novel
Pt22	TRAF3IP2 c.1366G>A p.Val456Met	Deleterious	Probably damaging		Novel
	FCGR3A g.161514759 delC ins CAG	-	-	18.07	1/114654
Pt23	PDHX c.171T>G p.Ile57Met	Deleterious	Probably damaging	24.0	Novel
Pt24	JAZF1 c.-23delA	-	-	18.15	16/109860
Pt25	MAP3K1 c.2_3insGGC p.Met1_Ala2insAla	-	-	18.68	Novel
Pt26	PHRF1 c.1355C>T p.Pro452Leu	Deleterious	Possibly damaging	23.9	Novel
Pt27	CD19 c.1318G>A p.Glu440Lys	Deleterious	Probably damaging	33	Novel
Pt28	CBLB c.621T>G p.Ile207Met	Deleterious	Probably damaging	23.9	Novel
	TMEM39A c.682_694delGCCTCGACTGTGG insA p.Ala228_Gly232delinsArg	-	-	20.7	Novel

Pt29	FCGR3A g.161514759 delC ins CAG TNXB	-	-	18.07	1/114654
Pt30	c.4535_4552delACGGGCAGCCCCAGG TGG p.Asp1512_Val1518del	-	-	32	9/110452
Pt31	PTPN22 c.314C>T p.Pro105Leu	Deleterious	Probably damaging	29.6	Novel
Pt32	PDHX c.171T>G p.Ile57Met	Deleterious	Probably damaging	24.0	Novel
